# COMPUTER SCIENCE CS3753
## Assignment #5

**Points**: 40 **Weight**: 2%
**Due**: Friday, November 2, 2018 at 11:55 pm in BlackBoard
**Note**: Late assignment will not be accepted without instructor's pre–approval.

Hand in a Jupyter notebook**, yourNmae-homework05.ipynb**, with appropriate markdown cells for descriptions and comments. Write Python code to solve each of the following questions. ***This homework must be completed individually.***
Include the following code at the beginning of your notebook and use it to plot figures. The arguments for the plotHistDist function are as follows. The func should be the pdf of a continuous random variable, x is a sequence of points on x-axis at which the probability density will be plotted, r is a set of values for which the histogram will be drawn. The last four parameters are all strings, for plot title, label, and labels for x and y axises. See lecture notes for examples.

```
import matplotlib.pyplot as plt
plt.rc('figure', figsize=(10, 6))

def plotDist(x, func, title, l, xlabel, ylabel) :
    # plot func for elements of x
    plt.plot(x, func, 'b-', lw=2, alpha=0.6, label=l)
    xl = plt.gca().get_xlim()
    #lines on Y-axis
    plt.hlines(0, xl[0], xl[1], linestyles='--', colors='#999999')
    plt.gca().set_xlim(xl)
    plt.legend(loc='best', frameon=False)
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.title(title)

def plotHistDist(func, x, r, title, l, xlabel, ylabel):
    plt.hist(r, normed=True, histtype='stepfilled', alpha=0.2)
    plotDist(x, func, title, l, xlabel, ylabel)
```

1. [25] In a earlier version of NHTSA Vehicle Tests database, stored in the CRASH.XLS file (provided on BlackBoard), the column DRIVHEAD measures the severity of a driver's head injury when the car is in a head-on collision with a fixed barrier while traveling at a 35 miles per hour.

   (a) Load the data in a DataFrame and perform any pre-processing as necessary.

   (b) Assume that the data population has a normal distribution, where the mean and standard deviation are the same as the mean and standard deviation of the data in the column.

Using plotHistDist function to plot both the histogram of the column and a normal distribution with the same mean and standard deviation in the same figure, and verify that the data is approximately normally distributed.

(c) A crash-tested car is selected at random and the head injury rating is observed.

i. Find the probability that the rating is between 500 and 700.
ii. Find the probability that the rating is between 400 and 500.
iii. Find the probability that the rating is less than 850.
iv. Find the probability that the rating is greater than 1000.
v. What rating will be exceeded by only 10% of tested cars?

2. [12] A department has 26 faculty members. For $i = 0, 1, 2, 3$, find $p_i$ the probability that $i$ of them were born on the Christmas Day. Assume that the birth rates are constant throughout the year and that each year has 365 days. Write Python code to find $p_i$assuming

(a) A binomial distribution.

(b) A Poisson distribution (which approximates the binomial distribution)